A Statistical Method for Fuzzy Time Series Forecasting Based on the Class Length in Tea Production

[1] R.V. Rosheba, [2]Dr.V. Anithakumari, [3]Dr. A. Gowri

[1]Research Scholar, Reg.No:23113272092003, Department of Mathematics, Vivekananda College, Agasteeswaram-629 701

[2]Assistant Professor, Department of Statistics, Government Arts and Science College, Nagercoil-929 004

[3]Assistant Professor, Department of Mathematics, Vivekananda College, Agasteeswaram-629 701

[1]roshebarv@gmail.com, [2]anithakumari_v @yahoo.co.in, [3]gowri.subburaj@yahoo.com

Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-27012, Tamilnadu, India.

ABSTRACT: Fuzzy forecasting time series is a method of foreseeing the anticipated data in situation in which a form of different time series is envisioned and the information is vague and imprecise. This study emphasis a new method which is used for fuzzy forecasting time series based on the class length and coefficient of variation is also used in the calculating forecasting values. The new method is examined on the tea production data of TATA Consumer Products Limited, Kerala. This new proposed technique is compared with the existing approaches to ascertain the efficacy in relations to mean square error (MSE) and the average forecasting error (AFE).

Keywords: Interval length, Fuzzy time series, Forecasting, Yule rule, MSE

1. Introduction

Tea production refers to the method of growing, harvesting, and processing tea leaves to produce diverse of tea such as black, green, white, and oolong. Tea production forecasting involves predicting future tea output based on historical data, weather conditions, and other influencing factors

Time series refers to a sequence of data points collected or recorded at specific time intervals. Time series analysis is mostly applied for forecasting future values based on past values. Fuzzy time series is a variation of time series analysis that incorporates fuzzy logic to handle uncertainty and vagueness in data. Fuzzy time series is often used for forecasting where the future values are predicted based on fuzzy relations derived from past data. Hwang (1998) recommended a technique according to the time variant in contract with the fuzzy predicting issues. Wang (2013) calculated to divide the universe of discourse into classes to progress forecasting value with the length which is not equal. Gangwar and Kumar (2012) projected a new model based on several partitioning to improve the accuracy in forecasted data. Chen (2017) proposed a forecasting method based on finest partitions of intervals in the universe of discourse and finest weighting. Pattanayak (2020) industrialized a new method using clustering c-means to govern the size of the class which is not equal.

This paper presents the statistical model constructed on class length for fuzzy forecasting time series in tea production. The proposed model and Yule method is used to partition the intervals.

The Mean Square Error (MSE) and the Average Forecasting Error (AFE) are calculated to find equivalence of these models.

2. Methodology

To determine the forecasting of Tea Production based on the interval length, we use two methods.

2.1 Yule's Method

In this method the class interval is classified using Yule's Rule. Yule's Rule suggests the following expression to arrive at approximate number of classes,

$$\boldsymbol{Y = 2.5 * (\sqrt[4]{t})}$$

Where Y is the number of classes and t is the total amount of data.

The algorithm to find the forecasting values are same as the new method.

2.2 Proposed Method

In this method, the class size is used to classify class length and the coefficient of variation is castoff to specify the time series fuzzy forecasting.

Step 1: The Universe of Discourse, u is determined,

$$C = E_{max} - E_{min}$$

$$h = 2\sqrt{n}$$

$$m = \frac{C}{h}$$

$$u = [E_{min}, E_{min} + m]$$

Step 2: Divide the Universe of Discourse into an equal number of intervals with equal length $u_{1,}u_2, \dots \dots . u_n$. The number of classes should be classified in to the number of linguistic variables $A_1, A_2, \dots \dots A_n$.

Step 3: linguistic variables A_i are framed and used the triangular membership rule to each class for creating fuzzy set.

Step 4: After the classification of intervals, the given values are fuzzified and the fuzzy logical relationships are applied. If A_i is the data of the year n and $n+1$, FLR is manifested as $A_i \rightarrow A_j$, where A_i is the present data and A_j is the following data.

Step 5: Given notations are shown as;

K_i is the actual data of the year n.

K_{i-1} is the actual data of the year $n-1$.

K_{i-2} is the actual data of the year $n-2$.

F_j is the approximate predicted data for the year $n+1$.

The expression for finding the predicting future value are as follows;

$$KD_i = |K_i - K_{i-1}|$$

$$KD_{ii} = |K_{i-1} - K_{i-2}|$$

$$\overline{KD} = \frac{KD_i + KD_{ii}}{2}$$

$$S_i = (KD_i + \overline{KD}) - (KD_{ii} + \overline{KD})$$

$$X_i = M[* A_j] + \frac{S_i * 100}{\overline{KD}}$$

$$F_j = X_i$$

3. Results and Discussion

3.1 Yule Method

Step 1: Finding the universe of discourse are as follows**,**

$$C = E_{max} - E_{min}$$

Here $E_{max} = 15203848$ and $E_{min} = 6951453$

$$C = 15203848 - 6951453$$

$$C = 8252395$$

By Yule's Rule, we find the number of intervals,

$$R = 2.5 * \left(\sqrt[4]{24}\right)$$

$$R = 7$$

Approximately, we have 7 intervals of equal length.

The size of the intervals,

$$h = \frac{8252395}{7} = 1178913.571 \cong 1178914$$

$u_1 = [6951453, 8130367]$, $u_2 = [8130367, 9309281]$, $u_3 = [9309281, 10488195]$,

$u_4 = [10488195, 11667109]$, $u_5 = [11667109, 12846023]$,

$u_6 = [12846023, 14024937]$, $u_7 = [14024937, 15203851]$.

The mid-point is obtained for each interval,

Table 1. Mid-points and Corresponding elements of Yule's Method

Interval	Boundaries	Midpoint	Corresponding element
u_1	[6951453, 8130367]	7540910	7781000, 7837166, 6951453, 7611811
u_2	[8130367, 9309281]	8719824	8451000, 9020549, 8240671, 8925957
u_3	[9309281, 10488195]	9898728	9536911, 10002663, 9718062, 9385546
u_4	[10488195, 11667109]	11077652	11008007, 11384121, 11604919, 11528599, 10976346
u_5	[11667109, 12846023]	12256566	11811670, 12424165, 12788294
u_6	[12846023, 14024937]	13435480	13299301, 13803725, 13965577
u_7	[14024937, 15203851]	14614394	15203848

Step 2:

The tea production data is classified into Seven intervals $(u_1, u_2, \dots\dots u_7)$. Therefore, total of 7 linguistic variables are described $(A_1, A_2, \dots\dots A_7)$.

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

$$A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

$$A_4 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0.5}{u_3} + \frac{1}{u_4} + \frac{0.5}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

$$A_5 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0.5}{u_4} + \frac{1}{u_5} + \frac{0.5}{u_6} + \frac{0}{u_7}$$

$$A_6 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7}$$

$$A_7 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0.5}{u_6} + \frac{1}{u_7}$$

Step 3: The given data is fuzzified.

If the tea production data of the year 2000 fits to the class u_2, then it is fuzzified to A_1. The actual and fuzzified tea production values are given in Table 2.

Table 2: Tea Production of Yule's Method of Fuzzy Set

Year	Actual Tea Production	Fuzzified production	Year	Tea Production	Fuzzified production
2000	8451000	A_2	2012	8240671	A_2
2001	7781000	A_1	2013	8925957	A_2
2002	7837166	A_1	2014	9536911	A_3
2003	11008007	A_4	2015	10002663	A_3
2004	11384121	A_4	2016	9718062	A_3
2005	11604919	A_4	2017	10976346	A_4
2006	11528599	A_4	2018	11811670	A_5
2007	13299301	A_6	2019	12424165	A_5
2008	15203848	A_7	2020	12788294	A_5
2009	9020549	A_2	2021	13803725	A_6
2010	6951453	A_1	2022	13965577	A_6
2011	7611611	A_1	2023	9385546	A_3

Step 4: The calculations for the forecasting results have been taken out by the given expressions of the previous above method. The results obtained are given in the Table 5.

Forecasting for 2003:

$$KD_i = |7837166 - 7781000| = 56166$$

$$KD_{ii} = |7781000 - 8451000| = 670000$$

$$\overline{KD} = \frac{56166 + 670000}{2} = 363083$$

$$S_i = |(670000 + 363083) - (56166 + 363083)| = 613834$$

$$X_{i=}11077652 + \frac{613834 * 100}{363083} = 11077821.06$$

$$F_j = 11077821$$

To find the exactness of fuzzy forecasting time series MSE and AFE are used. The MSE and AFE are calculated by the following expression;

$$\text{MSE} = \frac{\sum_{i=1}^{n} (Actual\,Value - \,Forecasted\,Value\,)^2}{n}$$

$$AFE(in\,\%) = \frac{Sum\,of\,\,forecasting\,errors}{n}$$

$$Forecasting\,erros(in\,\%) = \frac{|Actual\,value - Forecasted\,|}{Actual\,value} * 100$$

The forecasted results are given in Table 5.

3.2 Proposed Method

Step 1: For the given data, U is determined.

$E_{max} = 15203848$ and $E_{min} = 6951453$

$$C = 15203848 - 691453 = 8252395$$

$$h = 2\sqrt{24} = 9.79795 \cong 10$$

$$m = \frac{8252395}{10} = 825239.5 \cong 825240$$

$u_1 = [6951453, 7776693]$, $u_2 = [7776693, 8601933]$, $u_3 = [8601933, 9427173]$,

$u_4 = [9427173, 10252413]$, $u_5 = [10252413, 11077653]$,

$u_6 = [11077653, 11902893]$, $u_7 = [11902893,12728133]$,

$u_8 = [12728133, 13553373]$, $u_9 = [13553373, 14378613]$,

$u_{10} = [14378613, 15203853]$.

Table 3: Mid-points and Corresponding elements of Proposed Method

Interval	Boundaries	Midpoint	Corresponding element
u_1	[6951453,7776693]	7364073	6951453, 7611611
u_2	[7776693,8601933]	8189313	8451000, 7781000, 7837166,8240761
u_3	[8601933,9427173]	9014553	9020549, 8925957, 9385546
u_4	[9427173,10252413]	9839793	9536911, 10002663, 9718062
u_5	[10252413,11077653]	10665033	11008007, 10976346
u_6	[11077653,11902893]	11490273	11384121, 11604919, 11528599, 11811670
u_7	[11902893,12728133]	12315513	12424165
u_8	[12728133,13553373]	13140753	13299301, 12788294
u_9	[13553375,14378613]	13965993	13803725, 13965577
u_{10}	[14378613,15203853]	14791233	15203848

Step 2: For each interval linguistic variables are described.

The given data is spilt into ten classes $(u_1, u_2, \ldots\ldots u_{10})$, therefore ten linguistic variables $(A_1, A_2, \ldots\ldots A_{10})$ are described.

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_4 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0.5}{u_3} + \frac{1}{u_4} + \frac{0.5}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_5 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0.5}{u_4} + \frac{1}{u_5} + \frac{0.5}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_6 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7} + \frac{0}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_7 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0.5}{u_6} + \frac{1}{u_7} + \frac{0.5}{u_8} + \frac{0}{u_9} + \frac{0}{u_{10}}$$

$$A_8 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0.5}{u_7} + \frac{1}{u_8} + \frac{0.5}{u_9} + \frac{0}{u_{10}}$$

$$A_9 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0.5}{u_8} + \frac{1}{u_9} + \frac{0.5}{u_{10}}$$

$$A_{10} = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0.5}{u_9} + \frac{1}{u_{10}}$$

Step 3: The fuzzified data obtained as,

Table 4: Actual and Fuzzified Production of Proposed Method

Year	Actual Tea Production	Fuzzified production	Year	Tea Production	Fuzzified production
2000	8451000	A_2	2012	8240671	A_2
2001	7781000	A_2	2013	8925957	A_3
2002	7837166	A_2	2014	9536911	A_4
2003	11008007	A_5	2015	10002663	A_4
2004	11384121	A_6	2016	9718062	A_4
2005	11604919	A_6	2017	10976346	A_5
2006	11528599	A_6	2018	11811670	A_6
2007	13299301	A_8	2019	12424165	A_7
2008	15203848	A_{10}	2020	12788294	A_8

2009	9020549	A_3	2021	13803725	A_9
2010	6951453	A_1	2022	13965577	A_9
2011	7611611	A_1	2023	9385546	A_3

Step 4: Forecasting results.

Forecasting for 2003:

$$KD_i = |7837166 - 7781000| = 56166$$

$$KD_{ii} = |7781000 - 8451000| = 670000$$

$$\overline{KD} = \frac{56166 + 670000}{2} = 363083$$

$$S_i = |(670000 + 363083) - (56166 + 363083)| = 613834$$

$$X_{i=}10665033 + \frac{613834 * 100}{363083} = 10665202.06$$

$$F_j = 10665202$$

Yules and Proposed Method Forecasting value**,** Mean Square Error (MSE) and Average Forecasting Error (AFE) are evaluated and the values are given in below table.

Table 5: Forecasting results of Yule's and Proposed Method

Year	Actual Tea Production	Yule's Method	Proposed Method
2000	8451000	-	-
2001	7781000	-	-
2002	7837166	-	-
2003	11008007	11077821	10665202
2004	11384121	11077845	11490466
2005	11604919	11077810	11490431
2006	11528599	11077704	11490325
2007	13299301	13435577	13140802
2008	15203848	14614578	14791417
2009	9020549	8719828	9014557
2010	6951453	7541016	7364179
2011	7611611	7541010	7364173
2012	8240671	8719927	8189416
2013	8925957	8719829	9014558
2014	9536911	9898747	9839802
2015	10002663	9898750	9839805
2016	9718062	9898765	9839807
2017	10976346	11077700	10665081
2018	11811670	12256692	11490399
2019	12424165	12256606	12315553
2020	12788294	12256569	13140784
2021	13803725	13435531	13966044

2022	13965577	13435574	13966087
2023	9385546	9898883	9014698
MSE	-	126429293437	50575409691
AFE	-	**2.7653**	**1.6881**

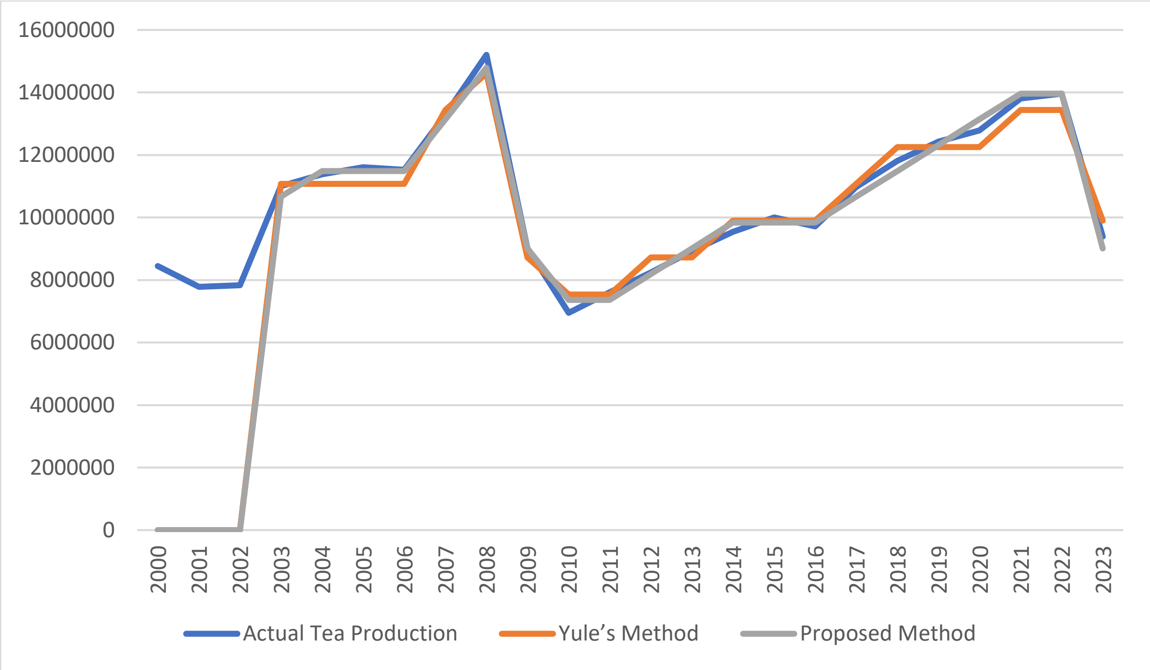

Figure 1: Comparison of Actual and Forecasted value of Yule's and Proposed Method

MSE and AFE values of Yule's and Proposed Method are compared.

Table 6: Comparison of MSE and AFE.

	Yule's Method	Proposed Method
MSE	126429293437	**50575409691**
AFE	2.7653	**1.6881**

In table 6, it shows that the Proposed Method is better than the Yule's Method by the compared value of MSE and AFE.

4.Conclusion

A Statistical structure is processed with better validity according to the class length. This Method has been used for predicting the fuzzy time series value of tea production and it is compared with the Yule's Method. MSE and AFE values are compared for the high accuracy. Based on the MSE and AFE values obtained by the both the methods, the Proposed Method provides the advanced precision than the Yule's Method. Therefore, the New Method is more desirable for predicting the future values using fuzzy forecasting time series.

References

1. Zadeh, L. A. (1965). Fuzzy sets, Information and control, 8(3): 338-353.

2. Song, Q. and Chissom, B.S. (1993): Forecasting enrollments with fuzzy time series part I, Fuczy sers and systems, 54(1): 1-9.
3. Hwang, J.R., Chen, S.M. and Lee, C.H. (1998). Handling forecasting problems using fuzzy time series, Fuzzy sets and systems, 100(1-3): 217-228.
4. Chen, S.M. and Tanuwijaya, K. (2011). Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques, Expert Systems with Applications, 38(12): 15425-15437.
5. Gangwar, S.S. and Kumar, S. (2012). Partitions based computational method for high-order fuzzy time series forecasting, Expert Systems with Applications, 39(15): 12158-12164.
6. Wang. L. Liu, X. and Pedrycz, W. (2013). Effective intervals determined by information granules to improve forecasting in fuzzy time series, Expert Systems with Applications, 40(14): 5673-5679.
7. Chen, S. M. and Phuong, B. D. H. (2017). Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors, Knowledge-Based Systems, 1 18: 204-216.
8. Pattanayak, R. M., Panigrahi, S. and Behera, H. S. (2020). High-order fuzzy time series forecasting by using membership values along with Data and Support Vector Machine, Arabian Journal for Science and Engineering, 45(12): 10311-10325.